

# Online Learning in Case of Unbounded Losses Using the Follow Perturbed Leader Algorithm\*

Vladimir V. V'yugin<sup>†</sup>

Institute for Information Transmission Problems, Russian Academy of Sciences,  
Bol'shoi Karetnyi per. 19, Moscow GSP-4, 127994, Russia  
e-mail vyugin@iitp.ru

August 26, 2010

## Abstract

In this paper the sequential prediction problem with expert advice is considered for the case where losses of experts suffered at each step cannot be bounded in advance. We present some modification of Kalai and Vempala algorithm of following the perturbed leader where weights depend on past losses of the experts. New notions of a volume and a scaled fluctuation of a game are introduced. We present a probabilistic algorithm protected from unrestrictedly large one-step losses. This algorithm has the optimal performance in the case when the scaled fluctuations of one-step losses of experts of the pool tend to zero.

**Keywords:** prediction with expert advice, follow the perturbed leader, unbounded losses, adaptive learning rate, expected bounds, Hannan consistency, online sequential prediction

## 1 Introduction

Experts algorithms are used for online prediction or repeated decision making or repeated game playing. Starting with the Weighted Majority Algorithm

---

\*This paper is an extended version of the ALT 2009 conference paper [19].

<sup>†</sup>This research was partially supported by Russian foundation for fundamental research: 09-07-00180-a and 09-01-00709a.

(WM) of Littlestone and Warmuth [11] and Vovk's [17] Aggregating Algorithm, the theory of Prediction with Expert Advice has rapidly developed in the recent times. Also, most authors have concentrated on predicting binary sequences and have used specific (usually convex) loss functions, like absolute loss, square and logarithmic loss. A survey can be found in the book of Lugosi, Cesa-Bianchi [12]. Arbitrary losses are less common, and, as a rule, they are supposed to be bounded in advance (see well known Hedge Algorithm of Freund and Shapire [6], Normal Hedge [2] and other algorithms).

In this paper, we consider a different general approach – “Follow the Perturbed Leader – FPL” algorithm, now called Hannan's algorithm [7], [10], [12]. Under this approach we only choose the decision that has fared the best in the past – the leader. In order to cope with adversary some randomization is implemented by adding a perturbation to the total loss prior to selecting the leader. The goal of the learner's algorithm is to perform almost as well as the best expert in hindsight in the long run. The resulting FPL algorithm has the same performance guarantees as WM-type algorithms for fixed learning rate and bounded one-step losses, save for a factor  $\sqrt{2}$ .

Prediction with Expert Advice considered in this paper proceeds as follows. We are asked to perform sequential actions at times  $t = 1, 2, \dots, T$ . At each time step  $t$ , experts  $i = 1, \dots, N$  receive results of their actions in form of their losses  $s_t^i$  - arbitrary real numbers.

At the beginning of the step  $t$  *Learner*, observing cumulating losses  $s_{1:t-1}^i = s_1^i + \dots + s_{t-1}^i$  of all experts  $i = 1, \dots, N$ , makes a decision to follow one of these experts, say Expert  $i$ . At the end of step  $t$  *Learner* receives the same loss  $s_t^i$  as Expert  $i$  at step  $t$  and suffers *Learner's* cumulative loss  $s_{1:t} = s_{1:t-1} + s_t^i$ .

In the traditional framework, we suppose that one-step losses of all experts are bounded, for example,  $0 \leq s_t^i \leq 1$  for all  $i$  and  $t$ .

Well known simple example of a game with two experts shows that Learner can perform much worse than each expert: let the current losses of two experts on steps  $t = 0, 1, \dots, 6$  be  $s_{0,1,2,3,4,5,6}^1 = (\frac{1}{2}, 0, 1, 0, 1, 0, 1)$  and  $s_{0,1,2,3,4,5,6}^2 = (0, 1, 0, 1, 0, 1, 0)$ . Evidently, the “Follow Leader” algorithm always chooses the wrong prediction.

When the experts one-step losses are bounded, this problem has been solved using randomization of the experts cumulative losses. The method of following the perturbed leader was discovered by Hannan [7]. Kalai and Vempala [10] rediscovered this method and published a simple proof of the

main result of Hannan. They called an algorithm of this type FPL (Following the Perturbed Leader).

The FPL algorithm outputs prediction of an expert  $i$  which minimizes

$$s_{1:t-1}^i - \frac{1}{\epsilon} \xi^i,$$

where  $\xi^i$ ,  $i = 1, \dots, N$ ,  $t = 1, 2, \dots$ , is a sequence of i.i.d random variables distributed according to the exponential distribution with the density  $p(x) = \exp\{-x\}$ , and  $\epsilon$  is a *learning rate*.

Kalai and Vempala [10] show that the expected cumulative loss of the FPL algorithm has the upper bound

$$E(s_{1:t}) \leq (1 + \epsilon) \min_{i=1, \dots, N} s_{1:t}^i + \frac{\log N}{\epsilon},$$

where  $\epsilon$  is a positive real number such that  $0 < \epsilon < 1$  is a learning rate,  $N$  is the number of experts.

Hutter and Poland [8], [9] presented a further developments of the FPL algorithm for countable class of experts, arbitrary weights and adaptive learning rate. Also, FPL algorithm is usually considered for bounded one-step losses:  $0 \leq s_t^i \leq 1$  for all  $i$  and  $t$ . Using a variable learning rate, an optimal upper bound was obtained in [9] :

$$E(s_{1:t}) \leq \min_{i=1, \dots, N} s_{1:t}^i + 2\sqrt{2T \ln N}.$$

Most papers on prediction with expert advice either consider bounded losses or assume the existence of a specific loss function (see [12]). We allow losses at any step to be unbounded. The notion of a specific loss function is not used.

The setting allowing unbounded one-step losses do not have wide coverage in literature; we can only refer reader to [1], [4], [14].

Poland and Hutter [14] have studied the games where one-step losses of all experts at each step  $t$  are bounded from above by an increasing sequence  $B_t$  given in advance. They presented a learning algorithm which is asymptotically consistent for  $B_t = t^{1/16}$ .

Allenberg et al. [1] have considered polynomially bounded one-step losses for a modified version of the Littlestone and Warmuth algorithm [11] under partial monitoring. In full information case, their algorithm has the expected

regret  $2\sqrt{N \ln N}(T+1)^{\frac{1}{2}(1+a+\beta)}$  in the case where one-step losses of all experts  $i = 1, 2, \dots, N$  at each step  $t$  have the bound  $(s_t^i)^2 \leq t^a$ , where  $a > 0$ , and  $\beta > 0$  is a parameter of the algorithm. They have proved that this algorithm is Hannan consistent if

$$\max_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^T (s_t^i)^2 < cT^a$$

for all  $T$ , where  $c > 0$  and  $0 < a < 1$ .

In this paper, we consider also the case where the loss grows “faster than polynomial, but slower than exponential”. A motivating example, where losses of the experts cannot be bounded in advance, is given in Section 4.

We present some modification of Kalai and Vempala [10] algorithm of following the perturbed leader (FPL) for the case of unrestrictedly large one-step expert losses  $s_t^i$  not bounded in advance:  $s_t^i \in (-\infty, +\infty)$ . This algorithm uses adaptive weights depending on past cumulative losses of the experts.

The full information case is considered in this paper. We analyze the asymptotic consistency of our algorithms using nonstandard scaling. We introduce new notions of *the volume of a game*  $v_t = v_0 + \sum_{j=1}^t \max_i |s_j^i|$  and *the scaled fluctuation* of the game  $\text{fluc}(t) = \Delta v_t / v_t$ , where  $\Delta v_t = v_t - v_{t-1}$  and  $v_0$  is a nonnegative constant.

We show in Theorem 1 that the algorithm of following the perturbed leader with adaptive weights constructed in Section 3 is asymptotically consistent in the mean in the case where  $v_t \rightarrow \infty$  and  $\Delta v_t = o(v_t)$  as  $t \rightarrow \infty$  with a computable bound. Specifically, if  $\text{fluc}(t) \leq \gamma(t)$  for all  $t$ , where  $\gamma(t)$  is a computable function such that  $\gamma(t) = o(1)$  as  $t \rightarrow \infty$ , our algorithm has the expected regret

$$2\sqrt{(6+\epsilon)(1+\ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t,$$

where  $\epsilon > 0$  is a parameter of the algorithm.

In case where all losses are nonnegative:  $s_t^i \in [0, +\infty)$ , we obtain a regret

$$2\sqrt{(2+\epsilon)(1+\ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t.$$

In particular, this algorithm is asymptotically consistent (in the mean) in a modified sense

$$\limsup_{T \rightarrow \infty} \frac{1}{v_T} E(s_{1:T} - \min_{i=1, \dots, N} s_{1:T}^i) \leq 0, \quad (1)$$

where  $s_{1:T}$  is the total loss of our algorithm on steps  $1, 2, \dots, T$ , and  $E(s_{1:T})$  is its expectation.

Proposition 1 of Section 2 shows that if the condition  $\Delta v_t = o(v_t)$  is violated the cumulative loss of any probabilistic prediction algorithm can be much more than the loss of the best expert of the pool.

In Section 3 we present some sufficient conditions under which our learning algorithm is Hannan consistent.<sup>1</sup>

In particular case, Corollary 1 of Theorem 1 says that our algorithm is asymptotically consistent (in the modified sense) in the case when one-step losses of all experts at each step  $t$  are bounded by  $t^a$ , where  $a$  is a positive real number. We prove this result under an extra assumption that the volume of the game grows slowly,  $\liminf_{t \rightarrow \infty} v_t/t^{a+\delta} > 0$ , where  $\delta > 0$  is arbitrary. Corollary 1 shows that our algorithm is also Hannan consistent when  $\delta > \frac{1}{2}$ .

At the end of Section 3 we consider some applications of our algorithm for the case of standard time-scaling.

In Section 4 we consider an application of our algorithm for constructing an arbitrage strategy in some game of buying and selling shares of some stock on financial market. We analyze this game in the decision theoretic online learning (DTOL) framework [6]. We introduce *Learner* that computes weighted average of different strategies with unbounded gains and losses. To change from the follow leader framework to DTOL we derandomize our FPL algorithm.

## 2 Games of prediction with expert advice with unbounded one-step losses

We consider a game of prediction with expert advice with arbitrary unbounded one-step losses. At each step  $t$  of the game, all  $N$  experts receive one-step losses  $s_t^i \in (-\infty, +\infty)$ ,  $i = 1, \dots, N$ , and the cumulative loss of the

---

<sup>1</sup> This means that (1) holds with probability 1, where  $E$  is omitted.

$i$ th expert after step  $t$  is equal to

$$s_{1:t}^i = s_{1:t-1}^i + s_t^i.$$

A probabilistic learning algorithm of choosing an expert outputs at any step  $t$  the probabilities  $P\{I_t = i\}$  of following the  $i$ th expert given the cumulative losses  $s_{1:t-1}^i$  of the experts  $i = 1, \dots, N$  in hindsight.

**Probabilistic algorithm of choosing an expert.**

FOR  $t = 1, \dots, T$

    Given past cumulative losses of the experts  $s_{1:t-1}^i$ ,  $i = 1, \dots, N$ , choose an expert  $i$  with probability  $P\{I_t = i\}$ .

    Receive the one-step losses at step  $t$  of the expert  $s_t^i$  and suffer one-step loss  $s_t = s_t^i$  of the master algorithm.

ENDFOR

The performance of this probabilistic algorithm is measured in its *expected regret*

$$E(s_{1:T} - \min_{i=1, \dots, N} s_{1:T}^i),$$

where the random variable  $s_{1:T}$  is the cumulative loss of the master algorithm,  $s_{1:T}^i$ ,  $i = 1, \dots, N$ , are the cumulative losses of the experts algorithms and  $E$  is the mathematical expectation (with respect to the probability distribution generated by probabilities  $P\{I_t = i\}$ ,  $i = 1, \dots, N$ , on the first  $T$  steps of the game).

In the case of bounded one-step expert losses,  $s_t^i \in [0, 1]$ , and a convex loss function, the well-known learning algorithms have expected regret  $O(\sqrt{T \log N})$  (see Lugosi, Cesa-Bianchi [12]).

A probabilistic algorithm is called *asymptotically consistent* in the mean if

$$\limsup_{T \rightarrow \infty} \frac{1}{T} E(s_{1:T} - \min_{i=1, \dots, N} s_{1:T}^i) \leq 0. \quad (2)$$

A probabilistic learning algorithm is called *Hannan consistent* if

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \left( s_{1:T} - \min_{i=1, \dots, N} s_{1:T}^i \right) \leq 0 \quad (3)$$

almost surely, where  $s_{1:T}$  is its random cumulative loss.

In this section we study the asymptotical consistency of probabilistic learning algorithms in the case of unbounded one-step losses.

Notice that when  $0 \leq s_t^i \leq 1$  all expert algorithms have total loss  $\leq T$  on first  $T$  steps. This is not true for the unbounded case, and there are no reasons to divide the expected regret (2) by  $T$ . We change the standard time scaling (2) and (3) on a new scaling based on a new notion of volume of a game. We modify the definition (2) of the normalized expected regret as follows. Define *the volume* of a game at step  $t$

$$v_t = v_0 + \sum_{j=1}^t \max_i |s_j^i|,$$

where  $v_0$  is a nonnegative constant. Evidently,  $v_{t-1} \leq v_t$  for all  $t$ .

A probabilistic learning algorithm is called *asymptotically consistent* in the mean (in the modified sense) in a game with  $N$  experts if

$$\limsup_{T \rightarrow \infty} \frac{1}{v_T} E(s_{1:T} - \min_{i=1, \dots, N} s_{1:T}^i) \leq 0. \quad (4)$$

A probabilistic algorithm is called Hannan consistent (in the modified sense) if

$$\limsup_{T \rightarrow \infty} \frac{1}{v_T} \left( s_{1:T} - \min_{i=1, \dots, N} s_{1:T}^i \right) \leq 0 \quad (5)$$

almost surely.

Notice that the notions of asymptotic consistency in the mean and Hannan consistency may be non-equivalent for unbounded one-step losses.

A game is called *non-degenerate* if  $v_t \rightarrow \infty$  as  $t \rightarrow \infty$ .

Denote  $\Delta v_t = v_t - v_{t-1}$ . The number

$$\text{fluc}(t) = \frac{\Delta v_t}{v_t} = \frac{\max_i |s_t^i|}{v_t}, \quad (6)$$

is called *scaled fluctuation* of the game at the step  $t$ .

By definition  $0 \leq \text{fluc}(t) \leq 1$  for all  $t$  (put  $0/0 = 0$ ).

The following simple proposition shows that each probabilistic learning algorithm is not asymptotically optimal in some game such that  $\text{fluc}(t) \not\rightarrow 0$  as  $t \rightarrow \infty$ . For simplicity, we consider the case of two experts and nonnegative losses.

**Proposition 1** *For any probabilistic algorithm of choosing an expert and for any  $\epsilon$  such that  $0 < \epsilon < 1$  two experts exist such that  $v_t \rightarrow \infty$  as  $t \rightarrow \infty$  and*

$$\begin{aligned} \text{fluc}(t) &\geq 1 - \epsilon, \\ \frac{1}{v_t} E(s_{1:t} - \min_{i=1,2} s_{1:t}^i) &\geq \frac{1}{2}(1 - \epsilon) \end{aligned}$$

for all  $t$ .

*Proof.* Given a probabilistic algorithm of choosing an expert and  $\epsilon$  such that  $0 < \epsilon < 1$ , define recursively one-step losses  $s_t^1$  and  $s_t^2$  of expert 1 and expert 2 at any step  $t = 1, 2, \dots$  as follows. By  $s_{1:t}^1$  and  $s_{1:t}^2$  denote the cumulative losses of these experts incurred at steps  $\leq t$ , let  $v_t$  be the corresponding volume, where  $t = 1, 2, \dots$

Define  $v_0 = 1$  and  $M_t = 4v_{t-1}/\epsilon$  for all  $t \geq 1$ . For  $t \geq 1$ , define  $s_t^1 = 0$  and  $s_t^2 = M_t$  if  $P\{I_t = 1\} \geq \frac{1}{2}$ , and define  $s_t^1 = M_t$  and  $s_t^2 = 0$  otherwise.

Let  $s_t$  be one-step loss of the master algorithm and  $s_{1:t}$  be its cumulative loss at step  $t \geq 1$ . We have

$$E(s_{1:t}) \geq E(s_t) = s_t^1 P\{I_t = 1\} + s_t^2 P\{I_t = 2\} \geq \frac{1}{2} M_t$$

for all  $t \geq 1$ . Also, since  $v_t = v_{t-1} + M_t = (1 + 4/\epsilon)v_{t-1}$  and  $\min_i s_{1:t}^i \leq v_{t-1}$ , the normalized expected regret of the master algorithm is bounded from below

$$\frac{1}{v_t} E(s_{1:t} - \min_i s_{1:t}^i) \geq \frac{2/\epsilon - 1}{1 + 4/\epsilon} \geq \frac{1}{2}(1 - \epsilon).$$

for all  $t$ . By definition

$$\text{fluc}(t) = \frac{M_t}{v_{t-1} + M_t} = \frac{1}{1 + \epsilon/4} \geq 1 - \epsilon$$

for all  $t$ .  $\triangle$

Proposition 1 shows that we should impose some restrictions of asymptotic behavior of  $\text{fluc}(t)$  to prove the asymptotic consistency of a probabilistic algorithm.



### 3 The Follow Perturbed Leader algorithm with adaptive weights

In this section we construct the FPL algorithm with adaptive weights protected from unbounded one-step losses.

Let  $\gamma(t)$  be a computable non-increasing real function such that  $0 < \gamma(t) < 1$  for all  $t$  and  $\gamma(t) \rightarrow 0$  as  $t \rightarrow \infty$ ; for example,  $\gamma(t) = 1/t^\delta$ , where  $\delta > 0$ . Let also  $a$  be a positive real number. Define

$$\alpha_t = \frac{1}{2} \left( 1 - \frac{\ln \frac{a(1+\ln N)}{2(e^{3/a}-1)}}{\ln \gamma(t)} \right) \text{ and} \quad (7)$$

$$\mu_t = a(\gamma(t))^{\alpha_t} = \sqrt{\frac{2a(e^{3/a}-1)}{(1+\ln N)}} (\gamma(t))^{1/2} \quad (8)$$

for all  $t$ , where  $e = 2.72 \dots$  is the base of the natural logarithm.<sup>2</sup>

Without loss of generality we suppose that  $\gamma(t) < \min\{A, A^{-1}\}$  for all  $t$ , where

$$A = \frac{2(e^{3/a}-1)}{a(1+\ln N)}.$$

We can obtain this choosing an appropriate value of the initial constant  $v_0$ . Then  $0 < \alpha_t < 1$  for all  $t$ .

We consider an FPL algorithm with a variable learning rate

$$\epsilon_t = \frac{1}{\mu_t v_{t-1}}, \quad (9)$$

where  $\mu_t$  is defined by (8) and the volume  $v_{t-1}$  depends on experts actions on steps  $< t$ . By definition  $v_t \geq v_{t-1}$  and  $\mu_t \leq \mu_{t-1}$  for  $t = 1, 2, \dots$ . Also, by definition  $\mu_t \rightarrow 0$  as  $t \rightarrow \infty$ .

Let  $\xi_t^1, \dots, \xi_t^N$ ,  $t = 1, 2, \dots$ , be a sequence of i.i.d random variables distributed according to the density  $p(x) = \exp\{-x\}$ . In what follows we omit the lower index  $t$ .

We suppose without loss of generality that  $s_0^i = v_0 = 0$  for all  $i$  and  $\epsilon_0 = \infty$ .

The FPL algorithm is defined as follows:

---

<sup>2</sup> The choice of the optimal value of  $\alpha_t$  will be explained later. It will be obtained by minimization of the corresponding member of the sum (42).

**FPL algorithm PROT.**

FOR  $t = 1, \dots, T$

Choose an expert with the minimal perturbed cumulated loss on steps  $< t$

$$I_t = \operatorname{argmin}_{i=1,2,\dots,N} \{s_{1:t-1}^i - \frac{1}{\epsilon_t} \xi^i\}. \quad (10)$$

Receive one-step losses  $s_t^i$  for experts  $i = 1, \dots, N$ , define  $v_t = v_{t-1} + \max_i s_t^i$  and  $\epsilon_{t+1}$  by (9).

Receive one-step loss  $s_t = s_t^{I_t}$  of the master algorithm.

ENDFOR

Let  $s_{1:T} = \sum_{t=1}^T s_t^{I_t}$  be the cumulative loss of the FPL algorithm on steps  $\leq T$ .

The following theorem shows that if the game is non-degenerate and  $\Delta v_t = o(v_t)$  as  $t \rightarrow \infty$  with a computable bound then the FPL-algorithm with variable learning rate (9) is asymptotically consistent.

We suppose that the experts are oblivious, i.e., they do not use in their work random actions of the learning algorithm. The inequality (12) of Theorem 1 below is reformulated and proved for non-oblivious experts at the end this section.

**Theorem 1** *Let  $\gamma(t)$  be a computable non-increasing real function such that  $0 \leq \gamma(t) \leq 1$  and*

$$\operatorname{fluc}(t) \leq \gamma(t) \quad (11)$$

*for all  $t$ . Then for any  $\epsilon > 0$  the expected cumulated loss of the FPL algorithm PROT with variable learning rate (9), where parameter  $a$  depends on  $\epsilon$ , is bounded:*

$$E(s_{1:T}) \leq \min_i s_{1:T}^i + 2\sqrt{(6 + \epsilon)(1 + \ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t \quad (12)$$

*for all  $t$ .*

*In case of nonnegative unbounded losses  $s_t^i \in [0, +\infty)$  we have a bound*

$$E(s_{1:T}) \leq \min_i s_{1:T}^i + 2\sqrt{(2 + \epsilon)(1 + \ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t. \quad (13)$$

Let also, the game be non-degenerate and  $\gamma(t) \rightarrow 0$  as  $t \rightarrow \infty$ . Then the algorithm PROT is asymptotically consistent in the mean

$$\limsup_{T \rightarrow \infty} \frac{1}{v_T} E(s_{1:T} - \min_{i=1, \dots, N} s_{1:T}^i) \leq 0. \quad (14)$$

*Proof.* The proof of this theorem follows the proof-scheme of [8] and [10].

Let  $\alpha_t$  be a sequence of real numbers defined by (7); recall that  $0 < \alpha_t < 1$  for all  $t$ .

The analysis of optimality of the FPL algorithm is based on an intermediate predictor IFPL (Infeasible FPL) with the learning rate  $\epsilon'_t$  defined by (15).

**IFPL algorithm.**

FOR  $t = 1, \dots, T$

Define the learning rate

$$\epsilon'_t = \frac{1}{\mu_t v_t}, \text{ where } \mu_t = a(\gamma(t))^{\alpha_t}, \quad (15)$$

$v_t$  is the volume of the game at step  $t$  and  $\alpha_t$  is defined by (7).

Choose an expert with the minimal perturbed cumulated loss on steps  $\leq t$

$$J_t = \operatorname{argmin}_{i=1,2,\dots,N} \{s_{1:t}^i - \frac{1}{\epsilon'_t} \xi^i\}.$$

Receive the one step loss  $s_t^{J_t}$  of the IFPL algorithm.

ENDFOR

The IFPL algorithm predicts under the knowledge of  $s_{1:t}^i$ ,  $i = 1, \dots, N$  (and  $v_t$ ), which may not be available at beginning of step  $t$ . Using unknown value of  $\epsilon'_t$  is the main distinctive feature of our version of IFPL.

For any  $t$ , we have  $I_t = \operatorname{argmin}_i \{s_{1:t-1}^i - \frac{1}{\epsilon'_t} \xi^i\}$  and  $J_t = \operatorname{argmin}_i \{s_{1:t}^i - \frac{1}{\epsilon'_t} \xi^i\} = \operatorname{argmin}_i \{s_{1:t-1}^i + s_t^i - \frac{1}{\epsilon'_t} \xi^i\}$ .

The expected one-step and cumulated losses of the FPL and IFPL algorithms at steps  $t$  and  $T$  are denoted

$$l_t = E(s_t^{I_t}) \text{ and } r_t = E(s_t^{J_t}),$$

$$l_{1:T} = \sum_{t=1}^T l_t \text{ and } r_{1:T} = \sum_{t=1}^T r_t,$$

respectively, where  $s_t^{J_t}$  is the one-step loss of the FPL algorithm at step  $t$  and  $s_t^{J_t}$  is the one-step loss of the IFPL algorithm, and  $E$  denotes the mathematical expectation.

**Lemma 1** *The cumulated expected losses of the FPL and IFPL algorithms with rearning rates defined by (9) and (15) satisfy the inequality*

$$l_{1:T} \leq r_{1:T} + 2(e^{3/a} - 1) \sum_{t=1}^T (\gamma(t))^{1-\alpha_t} \Delta v_t \quad (16)$$

for all  $T$ , where  $\alpha_t$  is defined by (7).

*Proof.* Let  $c_1, \dots, c_N$  be nonnegative real numbers and

$$m_j = \min_{i \neq j} \{s_{1:t-1}^i - \frac{1}{\epsilon_t} c_i\},$$

$$m'_j = \min_{i \neq j} \{s_{1:t}^i - \frac{1}{\epsilon'_t} c_i\} = \min_{i \neq j} \{s_{1:t-1}^i + s_t^i - \frac{1}{\epsilon'_t} c_i\}.$$

Let  $m_j = s_{1:t-1}^{j_1} - \frac{1}{\epsilon_t} c_{j_1}$  and  $m'_j = s_{1:t}^{j_2} - \frac{1}{\epsilon'_t} c_{j_2} = s_{1:t-1}^{j_2} + s_t^{j_2} - \frac{1}{\epsilon'_t} c_{j_2}$ . By definition and since  $j_2 \neq j$  we have

$$m_j = s_{1:t-1}^{j_1} - \frac{1}{\epsilon_t} c_{j_1} \leq s_{1:t-1}^{j_2} - \frac{1}{\epsilon_t} c_{j_2} \leq s_{1:t-1}^{j_2} + s_t^{j_2} - \frac{1}{\epsilon_t} c_{j_2} = \quad (17)$$

$$s_{1:t}^{j_2} - \frac{1}{\epsilon'_t} c_{j_2} + \left( \frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t} \right) c_{j_2} = m'_j + \left( \frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t} \right) c_{j_2}. \quad (18)$$

We compare conditional probabilities  $P\{I_t = j | \xi^i = c_i, i \neq j\}$  and  $P\{J_t = j | \xi^i = c_i, i \neq j\}$ .

The following chain of equalities and inequalities is valid:

$$\begin{aligned} & P\{I_t = j | \xi^i = c_i, i \neq j\} = \\ & P\{s_{1:t-1}^j - \frac{1}{\epsilon_t} \xi^j \leq m_j | \xi^i = c_i, i \neq j\} = \\ & P\{\xi^j \geq \epsilon_t (s_{1:t-1}^j - m_j) | \xi^i = c_i, i \neq j\} = \\ & P\{\xi^j \geq \epsilon'_t (s_{1:t-1}^j - m_j) + (\epsilon_t - \epsilon'_t) (s_{1:t-1}^j - m_j) | \xi^i = c_i, i \neq j\} \leq \end{aligned} \quad (19)$$

$$\begin{aligned} & P\{\xi^j \geq \epsilon'_t (s_{1:t-1}^j - m_j) + \\ & (\epsilon_t - \epsilon'_t) (s_{1:t-1}^j - s_{1:t-1}^{j_2} + \frac{1}{\epsilon_t} c_{j_2}) | \xi^i = c_i, i \neq j\} = \end{aligned} \quad (20)$$

$$\exp\{-(\epsilon_t - \epsilon'_t)(s_{1:t-1}^j - s_{1:t-1}^{j2})\} \times \quad (21)$$

$$P\{\xi^j \geq \epsilon'_t(s_{1:t-1}^j - m_j) + (\epsilon_t - \epsilon'_t)\frac{1}{\epsilon_t}c_{j2}|\xi^i = c_i, i \neq j\} \leq \quad (22)$$

$$\exp\{-(\epsilon_t - \epsilon'_t)(s_{1:t-1}^j - s_{1:t-1}^{j2})\} \times \\ P\{\xi^j \geq \epsilon'_t(s_{1:t}^j - s_t^j - m'_j - \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t}\right)c_{j2}) + \quad (23)$$

$$(\epsilon_t - \epsilon'_t)\frac{1}{\epsilon_t}c_{j2}|\xi^i = c_i, i \neq j\} = \quad (24)$$

$$\exp\{-(\epsilon_t - \epsilon'_t)(s_{1:t-1}^j - s_{1:t-1}^{j2}) + \epsilon'_t s_t^j\} \times \quad (25)$$

$$P\{\xi^j \geq \epsilon'_t(s_{1:t}^j - m'_j)|\xi^i = c_i, i \neq j\} =$$

$$\exp\left\{-\left(\frac{1}{\mu_t v_{t-1}} - \frac{1}{\mu_t v_t}\right)(s_{1:t-1}^j - s_{1:t-1}^{j2}) + \frac{s_t^j}{\mu_t v_t}\right\} \times \quad (26)$$

$$P\{\xi^j > \frac{1}{\mu_t v_t}(s_{1:t}^j - m'_j)|\xi^i = c_i, i \neq j\} \leq$$

$$\exp\left\{-\frac{\Delta v_t}{\mu_t v_t} \frac{(s_{1:t-1}^j - s_{1:t-1}^{j2})}{v_{t-1}} + \frac{\Delta v_t}{\mu_t v_t}\right\} \times \quad (27)$$

$$P\{\xi^j > \frac{1}{\mu_t v_t}(s_{1:t}^j - m'_j)|\xi^i = c_i, i \neq j\} =$$

$$\exp\left\{\frac{\Delta v_t}{\mu_t v_t} \left(1 - \frac{s_{1:t-1}^j - s_{1:t-1}^{j2}}{v_{t-1}}\right)\right\} P\{J_t = 1|\xi^i = c_i, i \neq j\}. \quad (28)$$

Here the inequality (19)-(20) follows from (17) and  $\epsilon_t \geq \epsilon'_t$ . We have used twice, in change from (20) to (21) and in change from (24) to (25), the equality  $P\{\xi > a + b\} = e^{-b}P\{\xi > a\}$  for any random variable  $\xi$  distributed according to the exponential law. The equality (22)-(23) follows from (18). We have used in change from (26) to (27) the equality  $v_t - v_{t-1} = \Delta v_t$  and the inequality  $|s_t^j| \leq \Delta v_t$  for all  $j$  and  $t$ .

The ratio in the exponent (28) is bounded :

$$\left|\frac{s_{1:t-1}^j - s_{1:t-1}^{j2}}{v_{t-1}}\right| \leq 2, \quad (29)$$

since  $\left|\frac{s_{1:t-1}^j}{v_{t-1}}\right| \leq 1$  for all  $t$  and  $i$ .

Therefore, we obtain

$$\begin{aligned}
& P\{I_t = j | \xi^i = c_i, i \neq j\} \leq \\
& \exp\left\{\frac{3}{\mu_t} \frac{\Delta v_t}{v_t}\right\} P\{J_t = j | \xi^i = c_i, i \neq j\} \leq \\
& \exp\{(3/a)(\gamma(t))^{1-\alpha_t}\} P\{J_t = j | \xi^i = c_i, i \neq j\}.
\end{aligned} \tag{30}$$

Since, the inequality (30) holds for all  $c_i$ , it also holds unconditionally

$$P\{I_t = j\} \leq \exp\{(3/a)(\gamma(t))^{1-\alpha_t}\} P\{J_t = j\}. \tag{31}$$

for all  $t = 1, 2, \dots$  and  $j = 1, \dots, N$ .

Since  $s_t^j + \Delta v_t \geq 0$  for all  $j$  and  $t$ , we obtain from (31)

$$\begin{aligned}
l_t + \Delta v_t &= E(s_t^{I_t} + \Delta v_t) = \sum_{j=1}^N (s_t^j + \Delta v_t) P(I_t = j) \leq \\
& \exp\{(3/a)(\gamma(t))^{1-\alpha_t}\} \sum_{j=1}^N (s_t^j + \Delta v_t) P(J_t = j) = \\
& \exp\{(3/a)(\gamma(t))^{1-\alpha_t}\} (E(s_t^{J_t}) + \Delta v_t) = \\
& \exp\{(3/a)(\gamma(t))^{1-\alpha_t}\} (r_t + \Delta v_t) \leq \\
& (1 + (e^{3/a} - 1)(\gamma(t))^{1-\alpha_t}) (r_t + \Delta v_t) = \\
& r_t + \Delta v_t + (e^{3/a} - 1)(\gamma(t))^{1-\alpha_t} (r_t + \Delta v_t) \leq \\
& r_t + \Delta v_t + 2(e^{3/a} - 1)(\gamma(t))^{1-\alpha_t} \Delta v_t.
\end{aligned} \tag{32}$$

In the last line of (32) we have used the inequality  $|r_t| \leq \Delta v_t$  for all  $t$  and the inequality  $\exp\{3r\} \leq 1 + (e^3 - 1)r$  for all  $0 \leq r \leq 1$ .

Subtracting  $\Delta v_t$  from both sides of the inequality (32) and summing it by  $t = 1, \dots, T$ , we obtain

$$l_{1:T} \leq r_{1:T} + 2(e^{3/a} - 1) \sum_{t=1}^T (\gamma(t))^{1-\alpha_t} \Delta v_t$$

for all  $T$ . Lemma 1 is proved.  $\triangle$

The following lemma, which is an analogue of the result from [10], gives a bound for the IFPL algorithm.

**Lemma 2** *The expected cumulative loss of the IFPL algorithm with the learning rate (15) is bounded :*

$$r_{1:T} \leq \min_i s_{1:T}^i + a(1 + \ln N) \sum_{t=1}^T (\gamma(t))^{\alpha_t} \Delta v_t \quad (33)$$

for all  $T$ , where  $\alpha_t$  is defined by (7).

*Proof.* The proof is along the line of the proof from Hutter and Poland [8] with an exception that now the sequence  $\epsilon'_t$  is not monotonic.

Let in this proof,  $\mathbf{s}_t = (s_t^1, \dots, s_t^N)$  be a vector of one-step losses and  $\mathbf{s}_{1:t} = (s_{1:t}^1, \dots, s_{1:t}^N)$  be a vector of cumulative losses of the experts algorithms. Also, let  $\xi = (\xi^1, \dots, \xi^N)$  be a vector whose coordinates are random variables.

Recall that  $\epsilon'_t = 1/(\mu_t v_t)$ ,  $\mu_t \leq \mu_{t-1}$  for all  $t$ , and  $v_0 = 0$ ,  $\epsilon'_0 = \infty$ .

Define  $\tilde{\mathbf{s}}_{1:t} = \mathbf{s}_{1:t} - \frac{1}{\epsilon'_t} \xi$  for  $t = 1, 2, \dots$ . Consider the vector of one-step losses  $\tilde{\mathbf{s}}_t = \mathbf{s}_t - \xi \left( \frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}} \right)$  for the moment.

For any vector  $\mathbf{s}$  and a unit vector  $\mathbf{d}$  denote

$$M(\mathbf{s}) = \operatorname{argmin}_{\mathbf{d} \in D} \{\mathbf{d} \cdot \mathbf{s}\},$$

where  $D = \{(0, \dots, 1), \dots, (1, \dots, 0)\}$  is the set of  $N$  unit vectors of dimension  $N$  and “ $\cdot$ ” is the inner product of two vectors.

We first show that

$$\sum_{t=1}^T M(\tilde{\mathbf{s}}_{1:t}) \cdot \tilde{\mathbf{s}}_t \leq M(\tilde{\mathbf{s}}_{1:T}) \cdot \tilde{\mathbf{s}}_{1:T}. \quad (34)$$

For  $T = 1$  this is obvious. For the induction step from  $T - 1$  to  $T$  we need to show that

$$M(\tilde{\mathbf{s}}_{1:T}) \cdot \tilde{\mathbf{s}}_T \leq M(\tilde{\mathbf{s}}_{1:T}) \cdot \tilde{\mathbf{s}}_{1:T} - M(\tilde{\mathbf{s}}_{1:T-1}) \cdot \tilde{\mathbf{s}}_{1:T-1}.$$

This follows from  $\tilde{\mathbf{s}}_{1:T} = \tilde{\mathbf{s}}_{1:T-1} + \tilde{\mathbf{s}}_T$  and

$$M(\tilde{\mathbf{s}}_{1:T}) \cdot \tilde{\mathbf{s}}_{1:T-1} \geq M(\tilde{\mathbf{s}}_{1:T-1}) \cdot \tilde{\mathbf{s}}_{1:T-1}.$$

We rewrite (34) as follows

$$\sum_{t=1}^T M(\tilde{\mathbf{s}}_{1:t}) \cdot \mathbf{s}_t \leq M(\tilde{\mathbf{s}}_{1:T}) \cdot \tilde{\mathbf{s}}_{1:T} + \sum_{t=1}^T M(\tilde{\mathbf{s}}_{1:t}) \cdot \xi \left( \frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}} \right). \quad (35)$$

By definition of  $M$  we have

$$M(\tilde{\mathbf{s}}_{1:T}) \cdot \tilde{\mathbf{s}}_{1:T} \leq M(\mathbf{s}_{1:T}) \cdot \left( \mathbf{s}_{1:T} - \frac{\xi}{\epsilon'_T} \right) = \min_{\mathbf{d} \in D} \{ \mathbf{d} \cdot \mathbf{s}_{1:T} \} - M(\mathbf{s}_{1:T}) \cdot \frac{\xi}{\epsilon'_T} . \quad (36)$$

The expectation of the last term in (36) is equal to  $\frac{1}{\epsilon'_T} = \mu_T v_T$ .

The second term of (35) can be rewritten

$$\begin{aligned} \sum_{t=1}^T M(\tilde{\mathbf{s}}_{1:t}) \cdot \xi \left( \frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}} \right) = \\ \sum_{t=1}^T (\mu_t v_t - \mu_{t-1} v_{t-1}) M(\tilde{\mathbf{s}}_{1:t}) \cdot \xi . \end{aligned} \quad (37)$$

We will use the inequality for mathematical expectation  $E$

$$0 \leq E(M(\tilde{\mathbf{s}}_{1:t}) \cdot \xi) \leq E(M(\xi) \cdot \xi) = E(\max_i \xi^i) \leq 1 + \ln N. \quad (38)$$

The proof of this inequality uses ideas of Lemma 1 from [8].

We have for the exponentially distributed random variables  $\xi^i$ ,  $i = 1, \dots, N$ ,

$$P\{\max_i \xi^i \geq a\} = P\{\exists i (\xi^i \geq a)\} \leq \sum_{i=1}^N P\{\xi^i \geq a\} = N \exp\{-a\}. \quad (39)$$

Since for any non-negative random variable  $\eta$ ,  $E(\eta) = \int_0^\infty P\{\eta \geq y\} dy$ , by (39) we have

$$\begin{aligned} E(\max_i \xi^i - \ln N) &= \int_0^\infty P\{\max_i \xi^i - \ln N \geq y\} dy \leq \\ &\int_0^\infty N \exp\{-y - \ln N\} dy = 1. \end{aligned}$$

Therefore,  $E(\max_i \xi^i) \leq 1 + \ln N$ .



By (38) the expectation of (37) has the upper bound

$$\sum_{t=1}^T E(M(\tilde{\mathbf{s}}_{1:t}) \cdot \xi)(\mu_t v_t - \mu_{t-1} v_{t-1}) \leq (1 + \ln N) \sum_{t=1}^T \mu_t \Delta v_t.$$

Here we have used the inequality  $\mu_t \leq \mu_{t-1}$  for all  $t$ ,

Since  $E(\xi^i) = 1$  for all  $i$ , the expectation of the last term in (36) is equal to

$$E\left(M(\mathbf{s}_{1:T}) \cdot \frac{\xi}{\epsilon'_T}\right) = \frac{1}{\epsilon'_T} = \mu_T v_T. \quad (40)$$

Combining the bounds (35)-(37) and (40), we obtain

$$\begin{aligned} r_{1:T} &= E\left(\sum_{t=1}^T M(\tilde{\mathbf{s}}_{1:t}) \cdot \mathbf{s}_t\right) \leq \\ &\min_i s_{1:T}^i - \mu_T v_T + (1 + \ln N) \sum_{t=1}^T \mu_t \Delta v_t \leq \\ &\min_i s_{1:T}^i + (1 + \ln N) \sum_{t=1}^T \mu_t \Delta v_t. \end{aligned} \quad (41)$$

Lemma is proved.  $\triangle$ .

We finish now the proof of the theorem.

The inequality (16) of Lemma 1 and the inequality (33) of Lemma 2 imply the inequality

$$\begin{aligned} E(s_{1:T}) &\leq \min_i s_{1:T}^i + \\ &+ \sum_{t=1}^T (2(e^{3/a} - 1)(\gamma(t))^{1-\alpha_t} + a(1 + \ln N)(\gamma(t))^{\alpha_t}) \Delta v_t. \end{aligned} \quad (42)$$

for all  $T$ .

The optimal value (7) of  $\alpha_t$  can be easily obtained by minimization of each member of the sum (42) by  $\alpha_t$ . In this case  $\mu_t$  is equal to (8) and (42) is equivalent to

$$E(s_{1:T}) \leq \min_i s_{1:T}^i + 2\sqrt{2a(e^{3/a} - 1)(1 + \ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t, \quad (43)$$

where  $a$  is a parameter of the algorithm PROT.

Also, for each  $\epsilon > 0$  an  $a$  exists such that  $2a(e^{3/a} - 1) < 6 + \epsilon$ . Therefore, we obtain (12).

We have  $\sum_{t=1}^T \Delta v_t = v_T$  for all  $T$ ,  $v_t \rightarrow \infty$  and  $\gamma(t) \rightarrow 0$  as  $t \rightarrow \infty$ . Then by Toeplitz lemma (see Lemma 4 of Section A)

$$\frac{1}{v_T} \left( 2\sqrt{(6 + \epsilon)(1 + \ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t \right) \rightarrow 0$$

as  $T \rightarrow \infty$ . Therefore, the FPL algorithm PROT is asymptotically consistent in the mean, i.e., the relation (14) of Theorem 1 is proved.  $\triangle$

In case where all losses are nonnegative:  $s_t^i \in [0, +\infty)$ , the inequality (29) can be replaced on

$$\left| \frac{s_{1:t-1}^j - s_{1:t-1}^{j_2}}{v_{t-1}} \right| \leq 1$$

for all  $t$  and  $i$ . In this case an analysis of the proof of Lemma 1 shows that the bound (43) can be replaced on

$$E(s_{1:T}) \leq \min_i s_{1:T}^i + 2\sqrt{a(e^{2/a} - 1)(1 + \ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t,$$

where  $a$  is a parameter of the algorithm PROT.

Since for each  $\epsilon > 0$  an  $a$  exists such that  $a(e^{2/a} - 1) < 2 + \epsilon$ , we obtain a version of (12) for nonnegative losses – the inequality (13).

We study now the Hannan consistency of our algorithm.

**Theorem 2** *Assume that all conditions of Theorem 2 hold and*

$$\sum_{t=1}^{\infty} (\gamma(t))^2 < \infty. \tag{44}$$

*Then the algorithm PROT is Hannan consistent:*

$$\limsup_{T \rightarrow \infty} \frac{1}{v_T} \left( s_{1:T} - \min_{i=1, \dots, N} s_{1:T}^i \right) \leq 0 \tag{45}$$

*almost surely.*

*Proof.* So far we assumed that perturbations  $\xi^1, \dots, \xi^N$  are sampled only once at time  $t = 0$ . This choice was favorable for the analysis. As it easily seen, under expectation this is equivalent to generating new perturbations  $\xi_t^1, \dots, \xi_t^N$  at each time step  $t$ ; also, we assume that all these perturbations are i.i.d for  $i = 1, \dots, N$  and  $t = 1, 2, \dots$ . Lemmas 1, 2 and Theorem 1 remain valid for this case. This method of perturbation is needed to prove the Hannan consistency of the algorithm PROT.

We use some version of the strong law of large numbers to prove the Hannan consistency of the algorithm PROT.

**Proposition 2** *Let  $g(x)$  be a positive nondecreasing real function such that  $x/g(x)$ ,  $g(x)/x^2$  are non-increasing for  $x > 0$  and  $g(x) = g(-x)$  for all  $x$ .*

*Let the assumptions of Theorem 1 hold and*

$$\sum_{t=1}^{\infty} \frac{g(\Delta v_t)}{g(v_t)} < \infty. \quad (46)$$

*Then the FPL algorithm PROT is Hannan consistent, i.e., (5) holds as  $T \rightarrow \infty$  almost surely.*

*Proof.* The proof is based on the following lemma.

**Lemma 3** *Let  $a_t$  be a nondecreasing sequence of real numbers such that  $a_t \rightarrow \infty$  as  $t \rightarrow \infty$  and  $X_t$  be a sequence of independent random variables such that  $E(X_t) = 0$ , for  $t = 1, 2, \dots$ . Let also,  $g(x)$  satisfies assumptions of Proposition 2. Then the inequality*

$$\sum_{t=1}^{\infty} \frac{E(g(X_t))}{g(a_t)} < \infty \quad (47)$$

*implies*

$$\frac{1}{a_T} \sum_{t=1}^T X_t \rightarrow 0 \quad (48)$$

*as  $T \rightarrow \infty$  almost surely.*

The proof of this lemma is given in Section A.

Put  $X_t = (s_t - E(s_t))/2$ , where  $s_t$  is the loss of the FPL algorithm PROT at step  $t$ , and  $a_t = v_t$  for all  $t$ . By definition  $|X_t| \leq \Delta v_t$  for all  $t$ . Then (47) is valid, and by (48)

$$\frac{1}{v_T}(s_{1:T} - E(s_{1:T})) = \frac{1}{v_T} \sum_{t=1}^T (s_t - E(s_t)) \rightarrow 0$$

as  $T \rightarrow \infty$  almost surely. This limit and the limit (14) imply (45).  $\triangle$

By Lemma 2 the algorithm PROT is Hannan consistent, since (44) implies (46) for  $g(x) = x^2$ . Theorem 2 is proved.  $\triangle$

Authors of [1] and [14] considered polynomially bounded one-step losses. We consider a specific example of the bound (42) for polynomial case.

**Corollary 1** *Assume that  $|s_t^i| \leq t^\alpha$  for all  $t$  and  $i = 1, \dots, N$ , and  $v_t \geq t^{\alpha+\delta}$  for all  $t$ , where  $\alpha$  and  $\delta$  are positive real numbers. Let also, in the algorithm PROT,  $\gamma(t) = t^{-\delta}$  and  $\mu_t = a(\gamma(t))^{\alpha_t}$ , where  $\alpha_t$  is defined by (7). Then*

- (i) *the algorithm PROT is asymptotically consistent in the mean for any  $\alpha > 0$  and  $\delta > 0$ ;*
- (ii) *this algorithm is Hannan consistent for any  $\alpha > 0$  and  $\delta > \frac{1}{2}$ ;*
- (iii) *the expected loss of this algorithm is bounded :*

$$E(s_{1:T}) \leq \min_i s_{1:T}^i + 2\sqrt{(6 + \epsilon)(1 + \ln N)} T^{1 - \frac{1}{2}\delta + \alpha} \quad (49)$$

*as  $T \rightarrow \infty$ , where  $\epsilon > 0$  is a parameter of the algorithm.*<sup>3</sup>

This corollary follows directly from Theorem 1, where condition (44) of Theorem 1 holds for  $\delta > \frac{1}{2}$ .

If  $\delta = 1$  the regret from (49) is asymptotically equivalent to the regret from Allenberg et al. [1] (see Section 1).

For  $\alpha = 0$  we have the case of bounded loss function ( $|s_t^i| \leq 1$  for all  $i$  and  $t$ ). The FPL algorithm PROT is asymptotically consistent in the mean if  $v_t \geq \beta(t)$  for all  $t$ , where  $\beta(t)$  is an arbitrary positive unbounded non-decreasing computable function (we can get  $\gamma(t) = 1/\beta(t)$  in this case). This algorithm is Hannan consistent if (44) holds, i.e.

$$\sum_{t=1}^{\infty} (\beta(t))^{-2} < \infty.$$

---

<sup>3</sup>Recall that given  $\epsilon$  we tune the parameter  $a$  of the algorithm PROT.

For example, this condition be satisfied for  $\beta(t) = t^{1/2} \ln t$ .

Theorem 1 is also valid for the standard time scaling, i.e., when  $v_T = T$  for all  $T$ , and when losses of experts are bounded, i.e.,  $\alpha = 0$ . Then for any  $\epsilon > 0$  the expected regret has the upper bound

$$2\sqrt{(6 + \epsilon)(1 + \ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \leq 4\sqrt{(6 + \epsilon)(1 + \ln N)T}$$

which is similar to bounds from [8] and [10].

Let us show that the bound (12) of Theorem 1 that holds against oblivious experts also holds against non-oblivious (adaptive) ones.

In non-oblivious case, it is natural to generate at each time step  $t$  of the algorithm PROT a new vector of perturbations  $\bar{\xi}_t = (\xi_t^1, \dots, \xi_t^N)$ ,  $\bar{\xi}_0$  is empty set. Also, it is assumed that all these perturbations are i.i.d according to the exponential distribution  $P$ , where  $i = 1, \dots, N$  and  $t = 1, 2, \dots$ . Denote  $\bar{\xi}_{1:t} = (\bar{\xi}_1, \dots, \bar{\xi}_t)$ .

Non-oblivious experts can react at each time step  $t$  on past decisions  $s_1, s_2, \dots, s_{t-1}$  of the FPL algorithm and on values of  $\bar{\xi}_1, \dots, \bar{\xi}_{t-1}$ .

Therefore, losses of experts and regret depend now from random perturbations:

$$\begin{aligned} s_t^i &= s_t^i(\bar{\xi}_{1:t-1}), \quad i = 1, \dots, N, \\ \Delta v_t &= \Delta v_t(\bar{\xi}_{1:t-1}), \end{aligned}$$

where  $t = 1, 2, \dots$

In non-oblivious case, condition (11) is a random event. We assume in Theorem 1 that in the game of prediction with expert advice regulated by the FPL-protocol the event

$$\text{fluc}(t) \leq \gamma(t) \text{ for all } t$$

holds almost surely.

An analysis of the proof of Theorem 1 shows that in non-oblivious case, the bound (12) is an inequality for the random variable

$$\begin{aligned} & \sum_{t=1}^T E(s_t) - \min_i s_{1:T}^i - \\ & - 2\sqrt{(6 + \epsilon)(1 + \ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t \leq 0, \end{aligned} \tag{50}$$

which holds almost surely with respect to the product distribution  $P^{t-1}$ , where the loss of the FPL algorithm  $s_t$  depend on a random perturbation  $\xi_t$  at step  $t$  and on losses of all experts on steps  $< t$ . Also,  $E$  is the expectation with respect to  $P$ .

Taking expectation  $E_{1:T-1}$  with respect to the product distribution  $P^{t-1}$  we obtain a version of (12) for non-oblivious case

$$E_{1:T} \left( s_{1:T} - \min_i s_{1:T}^i - 2\sqrt{(6+\epsilon)(1+\ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t \right) \leq 0$$

for all  $T$ .

## 4 An example: zero-sum experts

In this section we present an example of a game, where losses of experts cannot be bounded [20] in advance. Let  $S = S(t)$  be a function representing evolution of a stock price. Two experts will represent two concurrent methods of buying and selling shares of this stock.

Let  $M$  and  $T$  be positive integer numbers and let the time interval  $[0, T]$  be divided on a large number  $M$  of subintervals. Define a discrete time series of stock prices

$$S_0 = S(0), S_1 = S(T/(M)), S_2 = S(2T/(M)) \dots, S_M = S(T). \quad (51)$$

In this paper, volatility is an informal notion. We say that the difference  $(S_T - S_0)^2$  represents the macro volatility and the sum  $\sum_{i=0}^{T-1} (\Delta S_i)^2$ , where  $\Delta S_i = S_{i+1} - S_i$ ,  $i = 1, \dots, T-1$ , represents the micro volatility of the time series (51).

The game between an investor and the market looks as follows: the investor can use the long and short selling. At beginning of time step  $t$  Investor purchases the number  $C_t$  of shares of the stock by  $S_{t-1}$  each. At the end of trading period the market discloses the price  $S_{t+1}$  of the stock, and the investor incur his current income or loss  $s_t = C_t \Delta S_t$  at the period  $t$ . We have the following equality

$$(S_T - S_0)^2 = \left( \sum_{t=0}^{T-1} \Delta S_t \right)^2 =$$

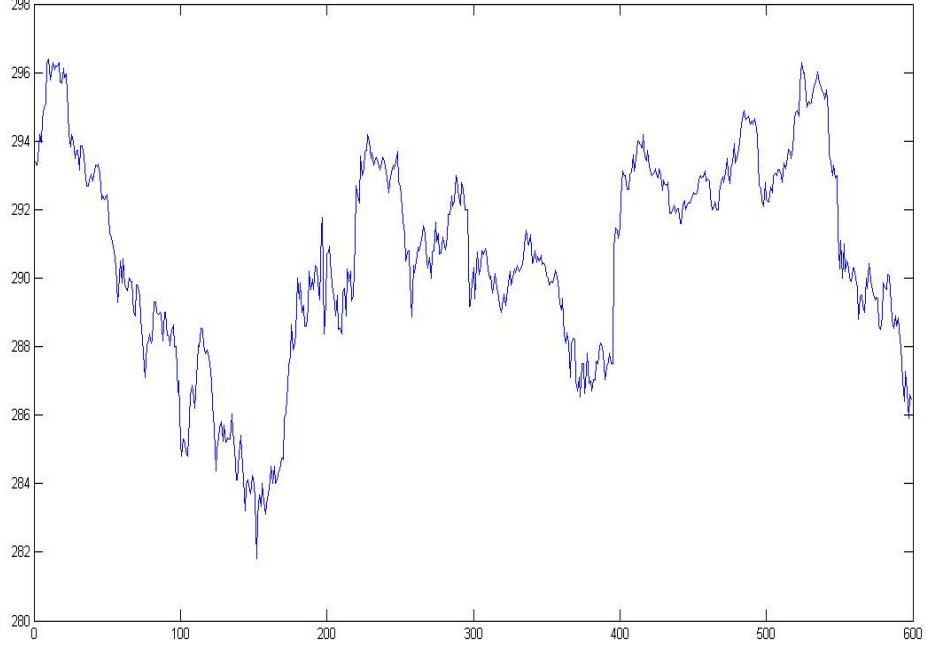


Fig. 1. Evolution of a stock price

$$= \sum_{t=0}^{T-1} 2(S_t - S_0)\Delta S_t + \sum_{t=0}^{T-1} (\Delta S_t)^2. \quad (52)$$

The equality (52) leads to the two strategies for investor which are represented by two experts. At the beginning of step  $t$  Experts 1 and 2 hold the number of shares

$$C_t^1 = 2C(S_t - S_0), \quad (53)$$

$$C_t^2 = -C_t^1, \quad (54)$$

where  $C$  is an arbitrary positive constant.

These strategies at step  $t$  earn the incomes  $s_t^1 = 2C(S_t - S_0)\Delta S_t$  and  $s_t^2 = -s_t^1$ . The strategy (53) earns in first  $T$  steps of the game the income

$$s_{1:T}^1 = \sum_{t=1}^T s_t^1 = 2C((S_T - S_0)^2 - \sum_{t=1}^{T-1} (\Delta S_t)^2).$$

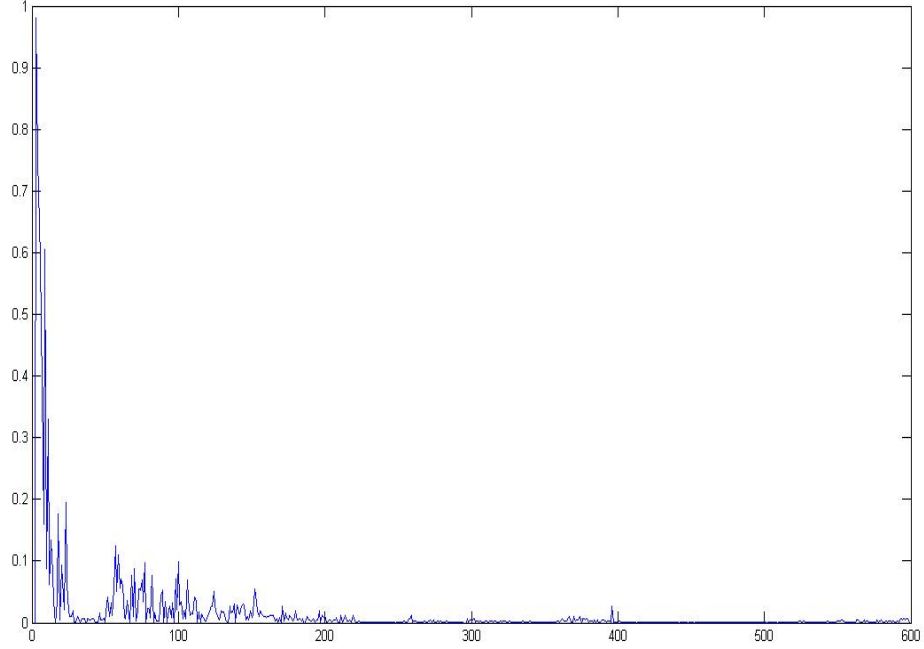


Fig. 2. Fluctuation of the game

The strategy (54) earns in first  $T$  steps the income  $s_{1:T}^2 = -s_{1:T}^1$ .

The number of shares  $C_t^1$  in the strategy (53) or number of shares  $C_t^2 = -C_t^1$  in the strategy (54) can be positive or negative. The one-step gains  $s_t^1$  and  $s_t^2 = -s_t^1$  are unbounded and can be positive or negative:  $s_t^i \in (-\infty, +\infty)$ .

Informally speaking, the first strategy will show a large return if  $(S_T - S_0)^2 \gg \sum_{i=0}^{T-1} (\Delta S_i)^2$ ; the second one will show a large return when  $(S_T - S_0)^2 \ll$

$\sum_{i=0}^{T-1} (\Delta S_i)^2$ . There is an uncertainty domain for these strategies, i.e., the case when both  $\gg$  and  $\ll$  do not hold. The idea of these strategies is based on the paper of Cheredito [3] (see also Rogers [15], Delbaen and Schachermayer [5]) who have constructed arbitrage strategies for a financial market that consists of money market account and a stock whose price follows a fractional



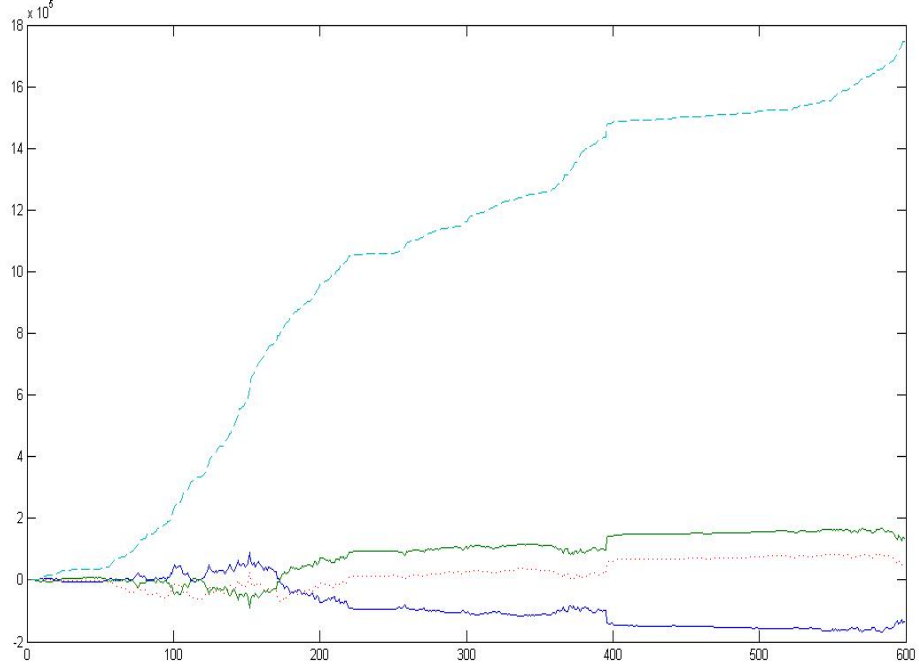


Fig. 3. Two symmetric solid lines – gains of two zero sums strategies, dotted line – expected gain of the algorithm PROT, dashed line – volume of the game

Brownian motion with drift or an exponential fractional Brownian motion with drift. Vovk [18] has reformulated these strategies for discrete time. We use these strategies to define a mixed strategy which incur gain when macro and micro volatilities of time series differ. There is no uncertainty domain for continuous time.

We analyze this game in the decision theoretic online learning (DTOL) framework [6]. We introduce *Learner* that can choose between two strategies (53) and (54). To change from the follow leader framework to DTOL we derandomize the FPL algorithm PROT.<sup>4</sup> We interpret the expected one-step gain  $E(s_t)$  gain as the weighted average of one-step gains of experts strategies. In more detail, at each step  $t$ , *Learner* divide his investment in proportion to the probabilities of expert  $t$  strategies (53) and (54) computed by the FPL

<sup>4</sup>To apply Theorem 1 we interpreted gain as a negative loss.

algorithm and suffers the gain

$$G_t = 2C(S_t - S_0)(P\{I_t = 1\} - P\{I_t = 2\})\Delta S_t$$

at any step  $t$ , where  $C$  is an arbitrary positive constant;  $G_{1:T} = \sum_{t=1}^T G_t = E(s_{1:T})$  is the *Learner's* cumulative gain.

Assume that  $|s_t^1| = o(\sum_{i=1}^t |s_i^1|)$  as  $t \rightarrow \infty$ . Let  $\gamma(t) = \mu$  for all  $t$ , where  $\mu$  is arbitrary small positive number. Then for any  $\epsilon > 0$

$$G_{1:T} \geq \left| \sum_{t=1}^T s_t^1 \right| - 2\mu^{1/2} \sqrt{(6 + \epsilon)(1 + \ln N)} \left( \sum_{t=1}^T |s_t^1| + v_0 \right)$$

for all sufficiently large  $T$ , and for some  $v_0 \geq 0$ .

Under condition of Theorem 1 we show that strategy of algorithm PROT is “defensive” in some weak sense :

$$G_{1:T} - \left| \sum_{t=1}^T s_t^1 \right| \geq -o \left( \sum_{t=1}^T |s_t^1| + v_0 \right)$$

as  $T \rightarrow \infty$ .

## 5 Conclusion

In this paper we try to extend methods of the theory of prediction with expert advice for the case when experts one-step gains cannot be bounded in advance. The traditional measures of performance do not work in general unbounded case. To measure the asymptotic performance of our algorithm, we replace the traditional time-scale on a volume-scale. New notion of volume of a game and scaled fluctuation of a game are introduced in this paper. In case of two zero-sum experts this notion corresponds to the sum of all transactions between experts.

Using the notion of the scaled fluctuation of a game, we can define very broad classes of games (experts) for which our algorithm PROT is asymptotically consistent in the modified sense. Also, restrictions on such games are formulated in relative terms: the logarithmic derivative of the volume of the game must be  $o(t)$  as  $t \rightarrow \infty$ .

A motivating example of a game with two zero-sum experts from Section 4 shows some practical significance of these problem. The FPL algorithm with

variable learning rates is simple to implement and it is bringing satisfactory experimental results when prices follow fractional Brownian motion.

There are some open problems for further research. It would be useful to analyze the performance of the well known algorithms from DTOL framework (like “Hedge” [6] or “Normal Hedge” [2]) for the case of unbounded losses in terms of the volume of a game.

There is a gap between Proposition 1 and Theorem 1, since we assume in this theorem that the game satisfies  $\text{fluc}(t) \leq \gamma(t) \rightarrow 0$ , where  $\gamma(t)$  is computable. Also, the function  $\gamma(t)$  is a parameter of our algorithm PROT. Does there exists an asymptotically consistent learning algorithm in case where  $\text{fluc}(t) \rightarrow 0$  as  $t \rightarrow \infty$  and where the function  $\gamma(t)$  is not a parameter of this algorithm?

A partial solution is based on applying “double trick” method to an increasing sequence of nonnegative functions  $\gamma_i(t)$  such that  $\gamma_i(t) \rightarrow 0$  as  $t \rightarrow \infty$  and  $\gamma_i(t) \leq \gamma_{i+1}(t)$  for all  $i$  and  $t$ . In this case a modified algorithm PROT is asymptotically consistent in the mean in any game such that

$$\limsup_{t \rightarrow \infty} \frac{\text{fluc}(t)}{\gamma_i(t)} < \infty$$

for some  $i$ .

We consider in this paper only the full information case. An analysis of these problems under partial monitoring is a subject for a further research.

## A Proof of Lemma 3

The proof of Lemma 3 is based on Kolmogorov’s theorem on three series and its corollaries. For completeness of presentation we reconstruct the proof from Petrov [13] (Chapter IX, Section 2).

For any random variable  $X$  and a positive number  $c$  denote

$$X^c = \begin{cases} X & \text{if } |X| \leq c \\ 0 & \text{otherwise.} \end{cases}$$

The Kolmogorov’s theorem on three series says:

For any sequence of independent random variables  $X_t$ ,  $t = 1, 2, \dots$ , the following implications hold

- If the series  $\sum_{t=1}^{\infty} X_t$  is convergent almost surely then the series  $\sum_{t=1}^{\infty} EX_t^c$ ,  $\sum_{t=1}^{\infty} DX_t^c$  and  $\sum_{t=1}^{\infty} P\{|X_t| \geq c\}$  are convergent for each  $c > 0$ , where  $E$  is the mathematical expectation and  $D$  is the variation.
- The series  $\sum_{t=1}^{\infty} X_t$  is convergent almost surely if all these series are convergent for some  $c > 0$ .

See Shiryaev [16] for the proof.

Assume conditions of Lemma 3 hold. We will prove that

$$\sum_{t=1}^{\infty} \frac{Eg(X_t)}{g(a_t)} < \infty \quad (55)$$

implies

$$\sum_{t=1}^{\infty} \frac{X_t}{a_t} < \infty$$

almost surely. From this, by Kroneker's lemma 5 (see below), the series

$$\frac{1}{a_t} \sum_{t=1}^{\infty} X_t \quad (56)$$

is convergent almost surely.

Let  $V_t$  be a distribution function of the random variable  $X_t$ . Since  $g$  non-increases,

$$P\{|X_t| > a_t\} \leq \int_{|x| \geq a_t} \frac{g(x)}{g(a_t)} dV_t(x) \leq \frac{Eg(X_t)}{g(a_t)}.$$

Then by (55)

$$\sum_{t=1}^{\infty} P\left\{\left|\frac{X_t}{a_t}\right| \geq 1\right\} < \infty \quad (57)$$

almost surely. Denote

$$Z_t = \begin{cases} X_t & \text{if } |X_t| \leq a_t \\ 0 & \text{otherwise.} \end{cases}$$

By definition  $x^2/g(x) \leq a_t/g(a_t)$  for  $|x| < a_t$ . Rearranging, we obtain  $x^2/a_t \leq g(x)/g(a_t)$  for these  $x$ . Therefore,

$$EZ_t^2 = \int_{|x| < a_t} x^2 dV_t(x) \leq \frac{a_t^2}{g(a_t)} \int_{|x| < a_t} g(x) dV_t(x) \leq \frac{a_t^2}{g(a_t)} Eg(X_t).$$

By (55) we obtain

$$\sum_{t=1}^{\infty} E \left( \frac{Z_t}{a_t} \right)^2 < \infty. \quad (58)$$

Since  $EX_t = \int_{-\infty}^{\infty} x dV_t(x) = 0$ ,

$$|EZ_t| = \left| \int_{|x| > a_t} x dV_t(x) \right| \leq \frac{a_t}{g(a_t)} \int_{|x| > a_t} g(x) dV_t(x) \leq \frac{a_t}{g(a_t)} Eg(X_t). \quad (59)$$

By (55)

$$\sum_{t=1}^{\infty} E \left( \frac{X_t}{a_t} \right)^1 \leq \sum_{t=1}^{\infty} \left| E \left( \frac{Z_t}{a_t} \right) \right| < \infty.$$

From (57)–(59) and the theorem on three series we obtain (56).

We have used Toeplitz and Kroneker's lemmas.

**Lemma 4** (*Toeplitz*) Let  $x_t$  be a sequence of real numbers and  $b_t$  be a sequence of nonnegative real numbers such that  $a_t = \sum_{i=1}^t b_i \rightarrow \infty$ ,  $x_t \rightarrow x$  and  $|x| < \infty$ . Then

$$\frac{1}{a_t} \sum_{i=1}^t b_i x_i \rightarrow x. \quad (60)$$

*Proof.* For any  $\epsilon > 0$  an  $t_\epsilon$  exists such that  $|x_t - x| < \epsilon$  for all  $t \geq t_\epsilon$ . Then

$$\left| \frac{1}{a_t} \sum_{i=1}^t b_i (x_i - x) \right| \leq \frac{1}{a_t} \sum_{i < t_\epsilon} |b_i (x_i - x)| + \epsilon$$

for all  $t \geq t_\epsilon$ . Since  $a_t \rightarrow \infty$ , we obtain (60).

**Lemma 5** (*Kroneker*) Assume  $\sum_{t=1}^{\infty} x_t < \infty$  and  $a_t \rightarrow \infty$ . Then  $\frac{1}{a_t} \sum_{i=1}^t a_i x_i \rightarrow 0$ .

The proof is the straightforward corollary of Toeplitz lemma.

## References

- [1] Allenberg C., Auer P., Györfi L., Ottucsak G. Hannan consistency in on-Line learning in case of unbounded losses under partial monitoring. In Proceedings of the 17th International Conference ALT 2006 (Jose L. Balcazar, Philip M. Long, Frank Stephan (Eds.)), LNCS, Volume 4264, 229-243, Springer-Verlag, Berlin, Heidelberg, 2006
- [2] Chaudhuri Kamalika, Freund Yoav, Hsu Daniel. A parameter-free hedging algorithm. Technical Report arXiv:0903.2851v1 [cs.LG], arXiv.org e-Print archive, March 2009.
- [3] Cheridito P. Arbitrage in fractional Brownian motion, Finance and Statistics, 7 (2003), 533-553
- [4] Cesa-Bianchi N., Mansour Y., Stoltz G. Improved second-order bounds for prediction with expert advice. Machine Learning, 66, Issue 2-3, 2007, 321-352
- [5] Delbaen F., Schachermayer W. A general version of the fundamental theorem of asset pricing. Mathematische Annalen, 300 (1994), 463-520
- [6] Freund Y, Schapire R.E. A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, 55 (1997), 119-139
- [7] Hannan J. Approximation to Bayes risk in repeated plays. In Contributions to the Theory of Games (M. Dresher, A.W. Tucker, and P. Wolfe, (Eds)), Volume 3, 97-139, Princeton University Press, 1957
- [8] Hutter M., Poland J. Prediction with expert advice by following the perturbed leader for general weights. In Proceedings of the 15th International Conference ALT 2004 (S.Ben-Dawid, J.Case, A.Maruoka (Eds.)), LNAI, Volume 3244, 279-293. Springer-Verlag, Berlin, Heidelberg, 2004
- [9] Hutter M., Poland J. Adeptive online prediction by following the Pperturbed leader, Journal of Machine Learning Research, 6, 2005, 639-660.
- [10] Kalai A., Vempala S. Efficient algorithms for online decisions. In Proceedings of the 16th Annual Conference on Learning Theory COLT

- 2003 (Bernhard Scholkopf, Manfred K. Warmuth (Eds.)), LNCS, Volume 2777, 506-521, Springer-Verlag, Berlin, 2003. Extended version in Journal of Computer and System Sciences, 71, 2005, 291-307
- [11] Littlestone N., Warmuth M.K. The weighted majority algorithm, Information and Computation, 108, 1994, 212-261
  - [12] Lugosi G., Cesa-Bianchi N. Prediction, Learning and Games. Cambridge University Press, New York, 2006
  - [13] Petrov V.V. Sums of independent random variables. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82, Springer-Verlag, New York, Heidelberg, Berlin, 1975
  - [14] Poland J., Hutter M. Defensive universal learning with experts. for general weight. In Proceedings of the 16th International Conference ALT 2005 (S.Jain, H.U.Simon and E.Tomita (Eds.)), LNAI, Volume 3734, 356-370. Springer-Verlag, Berlin, Heidelberg, 2005
  - [15] Rogers C. Arbitrage with fractional Brownian motion. Mathematical Finance, 7 (1997), 95-105
  - [16] Shiryaev A.N. Probability. Springer-Verlag, Berlin, 1980
  - [17] Vovk V.G. Aggregating strategies. In M. Fulk and J. Case, editors. Proceedings of the 3rd Annual Workshop on Computational Learning Theory, 371-383, San Mateo, CA, 1990. Morgan Kaufmann
  - [18] Vovk V. A game-theoretic explanation of the  $\sqrt{dt}$  effect, Working paper #5, 2003, Available online at <http://www.probabilityandfinance.com>
  - [19] V'yugin V.V. The Follow Perturbed Leader Algorithm Protected from Unbounded One-Step Losses. In Proceedings of the 20th International Conference ALT 2009 (R. Gavalda et al. (Eds.)): ALT 2009, LNAI 5809, pp. 38-52, Springer-Verlag Berlin Heidelberg 2009.
  - [20] V'yugin V.V. Learning Volatility of Discrete Time Series Using Prediction with Expert Advice. O. Watanabe and T. Zeugmann (Eds.): SAGA 2009, Lecture Notes in Computer Science 5792, pp. 16-30, Springer-Verlag Berlin Heidelberg 2009.